Lyon, G.R., & Flynn, J.M. (1991). Educational Validation Studies with Subtypes of Learnin Disabled Readers. In B.P. Rourke (Ed.), Neuropsychological Validation of Learning Disability Subtypes (pp. 223-242). New York: Guilford Press.

CHAPTER 11

Educational Validation Studies with Subtypes of Learning-Disabled Readers

G. Reid Lyon and Jane M. Flynn

A complicated task facing researchers, clinicians, and teachers is to identify and understand the instructional factors and decisions that must be considered when teaching learning-disabled (LD) individuals. By current definition, the learning disabled do not process information in a manner that allows them to comprehend, remember, and generalize concepts relevant to the development of reading, oral language, writing, mathematics, and/or social skills. Within this context, the value of treatment (e.g., educational remediation) is central to the concept of LD. From a clinical standpoint, it is our view that the descriptive and predictive validity of the LD diagnosis depends on the facility by which it generates testable hypotheses about instructional methodologies that have the highest probability of success with a given LD individual (Lyon, 1985a; Lyon & Moats, 1988; Lyon, Moats, & Flynn, 1988).

However, the complexity of studying instructional methods and outcomes with LD students is increased by factors related to the heterogeneous nature of the LD population, the multivariate and dynamic problem-solving demands inherent within the instructional decision-making process, and the methodological requirements necessary to identify valid interactions between "subtypes" of LD individuals and different forms of educational treatment. Each of these factors is considered here.

Population Heterogeneity and Treatment Effectiveness

It is well documented that individuals diagnosed as LD do not constitute a homogeneous group (Lyon & Watson, 1981; Rourke, 1985; Satz & Morris, 1981). In fact, current definitions of learning disabilities specify

that LD consists of several major subgroups identified on the basis of different handicapping conditions (e.g., oral language disorders, basic reading and reading comprehension disorders, arithmetic calculation and reasoning disorders, written language disorders). Recent classification research has also indicated that a number of these major diagnostic subgroups are themselves composed of homogeneous subtypes, each of which can be distinguished from one another by a particular array of information-processing, neuropsychological, and/or academic achievement characteristics. Space does not permit extensive discussion of these findings, but interested readers are referred to Fletcher and Morris (1986), Hooper and Willis (1989), and Rourke (1985) for extensive coverage of current LD subtype research findings.

The important point to note is that meaningful conclusions regarding treatment efficacy and outcomes for the learning disabled cannot be made until population heterogeneity is recognized, accounted for in an internally and externally valid fashion, and incorporated into LD subtype × treatment method designs. This point may take on added relevance when one considers that, even when samples of LD individuals are identified and grouped according to stringent selection criteria (e.g., similar academic achievement deficits, IQ above 100, middle to upper-middle socioeconomic status, and control for exposure to different curricula), there still exists substantial heterogeneity with respect to the development and manifestation of skills that are correlated with specific forms of academic underachievement. Thus, even within "well-defined" samples of LD individuals, there continues to exist substantial heterogeneity, indicating that not all LD children learn poorly for the same reasons; consequently, they do not respond equally well to the same teaching tactics or methodologies (see Lyon, 1985b).

Some advances have been made in accounting for both LD population and sample heterogeneity via the application of empirical, clinical, and rational classification methodologies to identify subtypes. However, our understanding of specific relationships between LD subtype characteristics and treatment responsiveness remains tentative because of a number of practical and logistical factors. We now turn to a discussion of some of these factors.

Factors That Impede the Educational Validation of LD Subtypes

Even when the heterogeneity of LD samples has been accounted for via the formation of typologies that contain individuals of like characteristics, understanding how and why each subtype responds to different

treatments is difficult for a variety of logistical and pedagogical reasons. For example, it is often the case that individuals who have participated in subtype identification studies are not available for subsequent treatment investigations. In other cases, parents and teachers are sometimes reluctant to have their children receive a form of instruction with which they may disagree or instruction that reduces time spent in other classes or school activities.

It also must be recognized that educational treatment outcomes are difficult to quantify accurately, priamrily because remediation may not produce immediate changes in cognitive and academic functioning. Moreover, the degree of relationship between subtype, instructional treatment method, and outcome is difficult to interpret because of limitations in accoutning for and measuring departures from treatment fidelity, teacher preparation and style, classroom climate, and the LD student's previous and concurrent instructional experiences (Lyon & Moats, 1988). Investigators planning to carry out educational validation studies with LD subtypes must not only consider the practical and logistical design factors addressed here, but they should also become acquainted with the theoretical and methodological features that exemplify quality aptitude (e.g., subtype) × treatment (e.g., teaching method) research.

Educational Validation of LD Subtypes: Theoretical and Methodological Considerations

In addition to the practical constraints cited earlier, another significant difficulty encountered when conducting educational validity studies of LD subtypes is satisfying the theoretical and methodological demands of classification research in general and aptitude × treatment experiments in particular. A necessary first step in such investigations is to delineate, in an *a priori* fashion, the scope, purpose, and theoretical basis for the subtype identification and educational validation study. From this context, theory-driven hypotheses can be formulated. These hypotheses, in turn, should lead to tentative descriptions of anticipated subtypes and specification of how predicted subtypes will respond to treatment.

Once theory-driven hypotheses have been developed to specify the possible range and nature of the subtype solution, both classification and external educational validation variables must be selected (see Lyon & Flynn, 1990, for a discussion of issues related to variable selection). Variables designated as classification variables are chosen to assess the critical attributes of the predicted subtypes. The educational validation variables (outcome measures) should be selected to assess multiple dimensions of change even within specific content domains (e.g., reading). This

is necessary because subtype × teaching method interactions may exist for some behaviors (e.g., reading accuracy) but not for others (e.g., reading speed). Variable selection is a complex process and should be guided by informed theoretical, psychometric, content, and developmental perspectives. As a general guideline for subtype validation studies, variables should be selected on the basis of the following:

1. Their theoretical coherence and ability to permit unconfounded and fine-grained analysis of hypothesized subtypal attributes and outcomes.
2. Their relationship to known paths of development within the content domain being studied (e.g., reading). This insures that the measurement of a subtypal attribute or validation domain is appropriate to the developmental level of the individuals in the sample.
3. Evidence that the variables constitute valid measurements of the classification (subtype) attributes and the educational validation treatments and outcomes.
4. Evidence that the variables possess adequate reliability.
5. Evidence that the classification and educational validation variables accomplish nonredundant assessments of subtypal attributes and treatment outcomes.

It is also useful to note that difference scores (e.g., scores reflecting change between pre- and posttest conditions) are frequently reported to be unreliable and thus limited in their usefulness in detecting treatment effects. Interested readers are referred to Cronbach and Snow (1977) for a discussion of this issue and recommendations for the selection of appropriate treatment-dependent variables.

The actual statistical procedures that can be used to detect subtype × teaching method interactions are complex and beyond the scope of this chapter. Although a brief overview of the recommended methodology is provided here, the reader is referred to Cronbach and Snow (1977) for more complete coverage. In general, studies designed to predict treatment response on the basis of subtype characteristics require regression analysis (Cronbach, 1977) and/or ANOVA designs (Bracht, 1970). In the case of the regression design, subjects are measured on a set of aptitude measures and then randomly assigned to one of two or more treatment (teaching) methods. Following completion of the teaching condition, the aptitude scores are used to predict the treatment outcome scores obtained by members of each group. The regression lines are then examined for parallelism and plotted to determine whether they interact within the effective range of the aptitude (subtype) measures.

In the ANOVA design, subjects are administered aptitude measures (e.g., neuropsychological tasks) and then assigned, according to their scores, to one of several levels of the aptitude (subtype) factor in a factorial analysis of variance. (Keep in mind that subtypes may be formed through multivariate empirical classification procedures, clinical assignment via visual inspection of the data, or application of rational grouping procedures.) Bracht (1970) has recommended that, following the teaching intervention(s), an ANOVA be computed to assess possible disordinal interactions among levels of the aptitude (subtype) factor and levels of the treatment (teaching method) factor. A disordinal interaction exists if the analyses of simple effects are significant *and* the mean cell differences have different algebraic signs.

When carrying out subtype × teaching method research, careful consideration has to be given to determining the number of subjects per aptitude × treatment condition needed to achieve a given level of power for the interaction. In the main, a relatively large sample size is recommended for this type of research. With both the regression and ANOVA designs, subjects should be equated on preintervention achievement and preintervention regression between achievement and aptitude *or* must be randomly assigned to the different teaching conditions if the results of the data analysis are to be interpretable. Obviously, the matching of subjects on preintervention achievement variables and the random assignment of subjects to teaching conditions would require a large sample size. This is particularly true if subjects from several identified subtypes are being randomly assigned to several teaching conditions.

No doubt, the complexity of conducting well-designed educational validation studies of LD subtypes has limited the number of investigations completed to date and has also reduced the quality of the few that have been published. Nevertheless, it may be useful to provide an overview of three research programs that are currently pursuing subtype remediation research. In doing so, our goal is to describe the current "state of the art" with an eye toward making improvements in future educational validation studies. The reader should note that Bakker's subtype validation studies are not reviewed in this chapter since he accomplishes this task in Chapter 7 (Bakker, Licht, & van Strien, this volume).

Selected Educational Validation Studies of LD Subtypes

To date, a few research programs have reported preliminary data that suggest that LD subtypes respond differently to various forms of treatment (remediation). Although the studies reported here have all been carried out with LD readers, the studies differ with respect to theoretical

orientation, assessment tasks used to characterize subtypes, and classification methodologies employed to identify subtypes.

For example, Lyon and his associates have identified several LD subtypes by applying empirical multivariate quantitative methods to information-processing task scores obtained by large samples of LD readers. External educational validation studies have then involved attempts to teach the disabled readers and to identify subtype × teaching method interactions. In contrast, Lovett and her colleagues (Lovett, Ramsby, & Barron, 1988) and Flynn and her group (Flynn & Deering, 1989) have concentrated on clinically identifying dyslexic subtypes on the basis of their reading and spelling patterns and then assigning subjects to different treatment conditions.

The three research groups have presented pilot data showing that children who display varied subtype attribute patterns respond in different ways to instructional formats. Each of these research programs is reviewed in greater detail in this section. Emphasis is placed on describing the theoretical orientation that serves as the context for variable selection, the types of treatment procedures used, and the clinical relationship between tasks and interventions. Attention is also given to the methodological shortcomings associated with each program of research.

The Lyon Research Program

Lyon and his associates (Lyon, 1983, 1985a, 1985b; Lyon, Stewart, & Freedman, 1982; Lyon & Watson, 1981) have questioned the appropriateness of a single-deficit classification model for reading disability and have hypothesized that LD readers (dyslexics) constitute a population that is composed of a number of subtypes, each of which is defined by its own particular array of linguistic, perceptual, and reading characteristics. The theoretical background underlying Lyon's research can be viewed as a logical extension of Luria's (1966, 1973) clinical neuropsychological theory and Benson and Geschwind's (1975) multiple-syndrome model of alexia. For example, Lyon (1983) proposed that reading development is a complex process that requires the concerted participation of cognitive, linguistic, and perceptual subskills. As such, deficiencies in any one subskill can limit the acquisition of fluent decoding and/or reading comprehesion abilities.

SUBTYPE VALIDATION STUDIES WITH OLDER LD READERS

Within this theoretical context, an initial series of studies was conducted (Lyon, Rietta, Watson, Porch, & Rhodes, 1981; Lyon & Watson, 1981) in

which a battery of tasks designed to assess linguistic and perceptual skills related to reading development was administered to 100 LD readers and 50 nomal readers matched for age (11–12 years) and IQ ($M = 104$). The data were submitted to a series of cluster analyses to test the hypothesis that subtypes could be identified. Six distinct subtypes were delineated and characterized by significantly different patterns of linguistic and perceptual deficits. The six-subtype solution remained stable across internal validation studies employing different variable subsets and clustering algorithms. Further, 94% of subjects were recovered into similar subtypes in a cross-validation study using a new subject sample (Lyon, 1983). A brief description of each of the subtypes' information-processing characteristics is provided here, followed by an overview of the intervention program. Readers are referred to cited references for specific details.

Children who were assigned empirically to Subtype 1 ($n = 10$) exhibited significant deficits in language comprehension, the ability to blend phonemes, visual–motor integration, visual–spatial skills, and visual memory skills, with strengths in naming and auditory discrimination skills. Analysis of the reading and spelling errors made by members of Subtype 1 indicated significant deficits in the development of both sight-word vocabulary and word-attack skills.

Children in Subtype 2 ($n = 12$) also exhibited a pattern of mixed deficits, but in a milder form than that observed in Subtype 1. Specifically, significant problems in language comprehension, auditory memory span, and visual–motor integration were observed and may have been related to the reading problems of these subjects. No deficits were seen in these youngsters' performance on naming, auditory discrimination, sound-blending, visual–spatial, and visual memory tasks. Subtype 2 members produced mixed orthographic and phonetic errors when reading, but to a much milder degree than did Subtype 1 children.

Members of Subtype 3 ($n = 12$) manifested selective deficits in language comprehension and sound blending, with corresponding strengths in all other linguistic and visual–perceptual skills measured. The oral reading errors made by Subtype 3 youngsters were primarily phonetic in nature, as would be expected from their diagnostic profile.

Children in Subtype 4 ($n = 32$) displayed significant deficiencies on a visual–motor integration task and average performance on all other measures. These youngsters presented with an assorted sample of oral reading errors, though most errors were made when they attempted to read phonetically irregular words.

Subtype 5 ($n = 12$) members displayed significant deficits in language comprehension, auditory memory span, and sound blending, with corresponding strengths in all measured visual–perceptual and visual–motor skills. These characteristics appeared related to the severity of their

oral reading and written spelling errors. The major academic characteristic that distinguished Subtype 5 youngsters from the other children was their consistently poor application of word-attack (phonetic) skills to the reading and spelling process.

The pattern of scores obtained by members of Subtype 6 ($n = 16$) indicated a normal diagnostic profile. These results were unexpected. It is quite possible that these children were reading poorly for reasons that were not detected by the assessment battery.

Following this subtype identification study, an external validation investigation (Lyon, 1983) was carried out to determine whether subtypes would respond differently to reading instruction. However, because of the relatively small sample size, an aptitude × treatment study designed according to the criteria discussed earlier could not be conducted. Therefore, it was decided to explore the possibility that the six subtypes might respond differently to one teaching condition. Since the children for this exploratory study had to be matched for preintervention achievement levels and other relevant variables (age, IQ, sex, socioeconomic status), the initial subject pool available from the subtype identification study was reduced to 30. Thus, random assignment of children from each of the six subtypes to several teaching conditions was not feasible.

In light of these logistical difficulties, five subjects were selected from each of the six subtypes. They were matched on their ability to read single words, age, IQ, race, and sex. All 30 subjects were white males ranging in age from 12.3 years to 12.7 years and in Full Scale IQ from 103.5 to 105. Preintervention grade equivalents on the Reading Recognition subtest of the Peabody Individual Achievement Test (PIAT; Dunn & Markwardt, 1970) ranged from 3.0 to 3.3, with centile ranks ranging from 4 to 8. It was not possible to control for the amount and type of previous reading instruction experienced by the children, their present curriculum, and the amount of time spent in classrooms for LD youngsters. Thus, the results obtained from this study must be evaluated in light of these confounding features.

The teaching method selected for the study was a synthetic phonics program (Traub & Bloom, 1975). This program was chosen because of its sequenced format, its coverage of major phonics concepts, and its familiarity to the teachers in training who were providing the instruction. All subjects were provided 1 hour of reading instruction per week (in addition to their special and regular classroom instruction) for 26 weeks.

Following the 26 hours of phonics instruction, the 30 children were posttested with the PIAT Reading Recognition subtest, and gain scores employing centile ranks were computed. A one-way analysis of variance indicated significant differences among the six subtypes for gain scores achieved from preintervention to posttesting. An analysis of subtype gain

scores and subsequent pairwise comparisons indicated that members of Subtype 6 made the most progress (mean centile rank gain = 18.0), followed by members of Subtype 4 (mean centile rank gain = 8.2). On the other hand, Subtypes 1, 2, 3, and 5 made minimal gains and were also not significantly different from one another in terms of the gains achieved. Subtypes 6 and 4 were both significantly different from one another and from all other subtypes with respect to their improvement in the oral reading of single words.

The data obtained in this subtype remediation study indicate that, for some subtypes, a synthetic phonics teaching intervention appeared to enhance significantly the ability to read single words accurately. Clearly, members of Subtypes 6 and 4 demonstrated robust improvements in their decoding capabilities. Whether or not the absence of auditory-verbal deficits in these two subtypes was associated with their good response to instruction cannot be answered clearly at this time, but one could hypothesize that this might be the case. This hypothesis is made more tenable by the observation that those subtypes with the most severe linguistic and memory-span deficits made either minimal gains (i.e., Subtypes 2 and 3) or no gains (i.e., Subtypes 1 and 5) in the ability to pronounce single words accurately and fluently.

SUBTYPE VALIDATION STUDIES WITH YOUNGER LD READERS

In a related program of research carried out with younger disabled readers (M age = 8.1 years) (Lyon, 1985b; Lyon et al., 1982), five LD subtypes were identified and validated internally and externally by using different variable subtests, clustering algorithms, and subtype × teaching method interaction studies. Again, a brief description of each of the subtypes' information-processing characteristics is provided, followed by an overview of the external validation intervention program.

Children assigned to Subtype 1 (n = 18) manifested significant deficits in visual perception, visual-spatial analysis and reasoning, and visual-motor integration. Visual memory was also below average, but not significantly so. All measured linguistic and verbal expressive skills were within the average range. The reading errors made by members of Subtype 1 appeared to be related to their diagnostic profile. Frequent mispronunciations resulting from confusion of orthographically similar words were noted, as were reading errors involving medial vowels and vowel combinations.

Children in Subtype 2 (n = 10) displayed selective deficits in morphosyntactic skills, sound blending, language comprehension, auditory memory span, auditory discrimination, and naming ability, with corresponding strengths in all measured visual-perceptual skills. These defic-

its across linguistic and verbal memory-span domains appeared to impede seriously their ability to decode single words and to apply decoding principles to the pronunciation of nonsense words.

Members of Subtype 3 ($n = 12$) scored in the normal range on all diagnostic measures and, thus, can be compared with subjects in the subtype identified by Lyon and Watson (1981) that scored significantly below normal on reading tasks without concomitant low performance on the diagnostic test battery. It is possible that members of Subtype 3 read inefficiently for social or affective reasons rather than because of inherent oral language or perceptual deficiencies. It is also quite possible that the diagnostic battery employed did not assess effectively all skills relevant to the developmental reading process. As was the case with Lyon and Watson's (1981) Subtype 6 (normal diagnostic profile), members of Subtype 3 scored higher than all other subgroups on the reading measures. These youngsters did have relatively more difficulties in comprehending reading passages than in the other measured reading skills. No systematic patterns of errors could be identified from analysis of their performance on word recognition and word attack measures.

Children in Subtype 4 ($n = 15$) displayed significant deficiencies in sound blending, language comprehension, auditory memory span, naming ability, and some aspects of visual perception. The difficulties manifested by Subtype 4 members in remembering, analyzing, synthesizing, and correctly sequencing verbal information appeared to have a significant effect on their ability to decode phonetically regular real and nonsense words.

Members of Subtype 5 ($n = 9$) manifested significant mixed deficits in morphosyntactic skills, sound blending, visual perception, visual-motor integration, visual-spatial analysis, and visual memory. These youngsters committed primarily orthographic errors when reading single words (both real and nonsense), possibly reflecting the influence of deficiencies in visual-verbal analysis and memory.

Following the subtype identification phase with the younger disabled readers, Lyon (1985b) carried out a pilot remediation study. Similar to the Lyon (1983) subtype remediation study, a relatively small sample size and other logistical difficulties (funding, sample migration) prohibited any attempts to assign members randomly from each of the five identified subtypes to a variety of teaching approaches. However, rather than teaching all subtype members with the same method, as was done in the first intervention study, one subtype (Subtype 2) was divided, with half of the members receiving reading instruction via a synthetic phonics approach and the other half receiving instruction through a combined whole-word and analytic phonics method.

Although this approach represents a significant departure from the experimental design necessary for an aptitude (subtype) × treatment (teaching method) interaction study, Lyon attempted to gain preliminary information about how children who are similar to one another diagnostically would respond to different teaching methods. Subtype 2 ($n = 10$) was chosen as the target subtype for this pilot study because all of its members displayed both significant linguistic deficits (auditory discrimination, auditory comprehension, auditory memory span) and verbal expressive deficits (retrieval, syntax, sequencing) within the context of robust visual perceptual-motor-memory strengths. Because all of the Subtype 2 members manifested significant difficulties reading single words and connected language, the opportunity existed to determine how two different reading approaches affected these skills in the presence of a number of linguistic subskill impairments.

For this pilot study, five children were randomly assigned to a synthetic phonics approach (Traub & Bloom, 1975), whereas the remaining five were placed randomly in a combined sight-word, contextual analysis, structural analysis, and analytic phonics group. Preintervention assessment using the Woodcock Reading Mastery Word Identification subtest (Woodcock, 1973) indicated that the five children in each remediation group were reading between the 8th and 10th centile ranks for age. The mean centile ranks for the two groups were not significantly different (Mann-Whitney $Z > .05$) prior to the initiation of the remediation programs.

Both remediation groups received approximately 30 hours of individualized instruction (3 hours a week for 10 weeks). Unfortunately, it was not possible to control for the type of previous exposure to reading instruction or for the type of ongoing regular and special class instruction the children were receiving in their typical school day. Thus, as in the Lyon (1983) study, any conclusions drawn from the results of this study must be interpreted in light of these confounding factors.

The synthetic phonics remediation group was taught via the scope and sequence presented in the Traub and Bloom (1975) reading program. A brief description of the instructional format for this approach was presented earlier. The combined remediation group learned to label whole words (three nouns, three verbs) rapidly by first pairing the words with pictures, then recognizing the names of the words (by a pointing response), and then finally reading the words in isolation. Following the development of rapid reading ability for these six words, function words (the, is, was, are) were introduced and taught. Following stable reading of these words, short sentences using combinations of the sight and function words were constructed and read in order to introduce the concept of

contextual analysis and to develop metalinguistic awareness of reading as a meaningful language skill. Following contextual reading drills, the combined group received instruction in structural analysis and the reading and comprehension of the morphosyntactical markers -ed, -s, and -ing. These morphemes were written as anagrams and introduced into context so that the children could readily grasp their effect on syntax and meaning. Finally, analytic phonics drills were initiated to develop letter–sound correspondences with the context of whole words. Specifically, phonetically regular words that could be read rapidly by sight were presented, and children were first asked to recognize a particular letter–sound correspondence ("Point to the letter that makes the /a/ sound.") and then to provide a recall response ("What sound(s) does this letter make?"). As children became more adept at recalling grapheme–phoneme relations, drills in auditory analysis and blending were initiated.

Following the 30 hours of remediation, children in both groups were posttested with an alternate form of the Woodcock Reading Mastery Word Identification subtest. Significant differences were found between the two remediation groups with respect to postintervention reading centile rank scores (Mann–Whitney $A < .003$). Children within the combined remediation group gained, on the average, 11 centile rank points, whereas members of the synthetic phonics group gained approximately 1 centile rank point.

There is little doubt that Subtype 2 members responded significantly differently to two forms of reading instruction. Apparently, the auditory receptive and auditory expressive language deficits that characterized each member of Subtype 2 impeded response to a reading instructional method that required learning letter–sound correspondences in isolation followed by blending and contextual reading components. A tentative hypothesis might be that Subtype 2 children did not have the linguistic subskills necessary for success with this approach but could deploy their relatively robust visual–perceptual and memory skills more effectively with whole words, as seen within the combined remediation. A more tenable hypothesis is that whole-word reading placed far less linguistic demands on these readers than did alphabetic approaches that require a phonological awareness of (1) sound structure and acoustic boundaries and (2) the relationship of these units to letter sequences. Thus, whereas Subtype 2 members learned to read whole words in structured, isolated context, their ability to generalize phonological concepts to read new words remained limited.

In general, the data derived from this series of subtype identification and remediation studies support a model of dyslexia that presumes that a number of diverse information-processing deficits can have specific read-

ing disability as a common correlate. Although the results from these basic research endeavors are interesting, the findings have limited clinical utility for a number of reasons. First, the kinds of subtypes identified and their descriptions are limited by the range and quality of the classification tasks that provided the data for cluster analysis. For example, the tasks selected for use in Lyon's assessment batteries did not provide adequate fine-grained coverage of some linguistic factors (particularly phonology) implicated in the developmental reading process. Second, the specific nature of the relationship between subtype assessment characteristics and response to reading instruction is difficult to determine because the assessment tasks are indirect measures of associated symptomatology. Third, it is not well understood whether the correlated information-processing deficits constitute necessary and/or sufficient conditions for reading disability. Fourth, methodological limitations in sample size and the number and type of dependent reading measures preclude adequate interpretations and confident generalization of the subtype × teaching method interaction studies. Finally, even though particular teaching (treatment) approaches had differential effects for some subtypes, it is difficult to determine if the effects should be attributed to subtype characteristics, the instructional program, the interaction between the two, the teacher, the time spent in remediation, or previous or concomitant educational experiences.

The Lovett Research Program

A subtype × teaching method study by Lovett et al. (1988) exemplifies a classification system that used direct measures of reading behaviors to identify subtypes of reading-disabled children who fail at crucial stages of reading acquisition. Lovett and her colleagues hypothesized that some readers (accuracy-disabled) fail at the initial stage of reading development, which involves the ability to recognize accurately and decode single words. Others (rate-disabled), although accurate decoders, demonstrate a deficiency at a later stage of reading characterized by the development of speed and fluency. Differential response to treatment was predicted on the basis of pretreatment reading and language characteristics of children assigned to either an accuracy-disabled or a rate-disabled subtype.

From a clinic-referred sample of 8- to 13-year-old children, Lovett selected 112 disabled readers based on multiple measures of accuracy and speed of single-word recognition and contextual reading. Accuracy-disabled readers were identified on the basis of substandard performance on four out of five tests of untimed word recognition. Rate-disabled children were classified on similar criteria for reading speed in the presence of

average decoding accuracy. The groups did not differ significantly in age or IQ. However, accuracy-disabled children were significantly different from rate-disabled subjects on pretreatment language characteristics (i.e., receptive vocabulary, naming opposites, and syntax).

Because it is generally desirable to classify children based on nonredundant measurements (see Lyon & Flynn, 1990), a word about Lovett's use of multiple measurements to assess single constructs is in order. Because standardized reading tests vary significantly in the number, regularity, and complexity of words used as stimuli (Lovett et al., 1988; Voeller & Armus, 1988), Lovett used multiple measures of critical reading skills in order to avoid a classification artifact based on the idiosyncratic content of single instruments.

Following identification, children from each subtype received 40 hours of remediation with random assignment to one of two treatments or a control condition. A decoding skills program (DS) emphasized acquisition of single-word recognition for both regular and exception words of high frequency. The second treatment condition (OWLS) emphasized reading in context, listening and reading comprehension, vocabulary development, syntactical elaboration, and written compositions. Finally, a control condition taught classroom survival skills (CSS) with no direct instruction in reading or exposure to print.

Pre- and posttests measured several dimensions of reading skill. The dependent variables included an experimental test of phonetically regular and exception words, standardized tests of word recognition singly and in context, and language measures. Given the theoretical framework of the study, these measures appear to constitute ecologically valid measurements of subtypal attributes and treatment conditions. Unfortunately, reliability data were not reported. This information would have been additionally helpful in interpreting the results of the experimental word recognition test, which does appear to have adequate content validity from both a classification and treatment standpoint.

Data were analyzed using a series of ANOVAs. For instance, analysis of the experimental word task included a five-way analysis of variance with nesting of subjects according to sample (rate-disabled, accuracy-disabled) and treatment program (DS, OWLS, CSS). Repeated measures factors consisted of word type (regular, exception), word frequency class (high, medium, low), and time of test (pretest, posttest). Posttest cell means were adjusted for pretest performance differences.

Results of the experimental word recognition test were interesting with respect to subtype × teaching method interactions. For example, rate-disabled readers in both the DS and OWLS conditions made significant gains on nonphonetic words but not on phonetically regular words as compared to their control peers in CSS. On the other hand, accuracy-

disabled readers demonstrated posttest gains for both regular and nonphonetic words, with greater gains made on nonphonetic words in the DS condition. This latter finding was unexpected, given the presumed additional processing demands for remembering sight words as opposed to words that conform to sound–symbol patterns. Lovett et al. (1988) speculated that this finding reflected the fact that more time was spent teaching individual nonphonetic words, while phonetically regular words were presented in word patterns. An alternate explanation could be that nonphonetic words represented more salient entities than phonetically regular words presented in pattern drills and were thus more readily mastered. The ambiguity inherent in this interpretation illustrates the complexity of accounting for educational outcomes even in well-controlled treatments with relatively precise dependent variables.

The standardized test results also illustrate the difficulty of measuring change given a relatively brief treatment period. Both accuracy-disabled and rate-disabled readers in DS and OWLS demonstrated significantly greater performance on the Wide Range Achievement Test—Revised (Jastak & Wilkinson, 1984) compared to children in CSS. However, inspection of the posttest raw data reveals that superior posttest performance involved raw score differences of only one or two words. Because CSS children received no reading instruction, this statistically significant finding does not appear to be clinically useful in choosing a treatment program.

Despite the difficulties noted, the Lovett et al. (1988) subtype × teaching method study is important for a number of reasons. First, the classification system was based on direct observation of reading behaviors within a developmental framework. Second, multiple measures were used to classify children, thus avoiding classification artifacts associated with sizable differences in word type found in different standardized tests. Third, a large number of disabled readers, well matched on IQ and age, were assigned to clearly defined reading treatments and a nonreading control condition. Fourth, project teachers implemented all three treatments, thus controlling for teacher differences. The distinctiveness and fidelity of each treatment were also well documented through random observation and coding of student and teacher activities by two raters with resultant good interrater reliability. Finally, the dependent measures included theory-based experimental tasks as well as standardized tests of reading skills in order to assess educational outcomes adequately.

The Flynn Research Program

A similar classification and educational validation pilot study of longer duration was conducted by Flynn (see Flynn & Deering, 1989; Lyon et al.,

1988). Similar to the Lovett et al. (1988) study, direct measures of reading and spelling behaviors were chosen to classify children and to generate hypotheses about instructional methodologies that have a predicted probability of success with identified subtypes.

Assessment tasks were chosen to represent key subskills in an interactive reading model conceptualized to include lower-level processes (alphabetic, logographic, orthographic word recognition skills) and higher-level processes (syntax, semantics, experiential knowledge, executive monitoring skills). Ecological assessment of intact, deficient, and compensatory reading strategies included oral reading samples to classify errors and the Boder Tests of Reading–Spelling Behaviors (Boder & Jarrico, 1982). Thus, test content was directly related to the tasks that children face on a daily basis and led to testable hypotheses regarding subtype responses to specified instructional methodologies.

Children who demonstrated significant difficulties with sound–symbol relationships, the decoding and spelling of single words, and contextual reading were classified as dysphonetic. These children are clinically similar to Lovett's accuracy-disabled readers. Dyseidetic (rate-disabled) children demonstrated normal decoding abilities on phonetically regular words and produced good phonetic equivalents in their misspellings, but they had difficulty recognizing nonphonetic words and read slowly.

Using these criteria, Flynn identified 27 first, second, and third grade children. Because of limited sample size, matching on IQ, age, and other variables generally related to achievement was not possible. The two groups did not differ significantly with respect to receptive vocabulary, age, parents' educational level, and school history. This pilot study was conducted during the school day, and therefore random assignment to treatment was difficult because of classroom scheduling conflicts. Nevertheless, children's assignments to treatment were randomized as much as possible across ages and subtypes. As in the Lovett et al. (1988) study, no attempt was made to control for concomitant reading instruction in the regular classroom or at home. The children participated in reading remediation for 33 weeks, three times a week for 45 minutes per session during the first year of the study and for four sessions per week during the second year. Differences in teacher preparation and style were accounted for by assigning teachers to different remediation groups at midyear.

It was hypothesized that dysphoneic (accuracy-disabled) readers would respond to a program that controlled the phonetic complexity of English orthography through use of the Initial Teaching Alphabet (ITA). In addition, phonetic principles were presented analytically to children in this subtype through a language experience approach. Conversely, dyseidetic (rate-disabled) children were hypothesized to have in-

tact auditory-linguistic reading skills and were predicted to respond best to a synthetic phonics program. The Distar Reading Program (Engelmann & Bruner, 1984) was chosen as the synthetic phonics program. Children in the two treatment groups did not differ significantly on age, verbal IQ, or pretreatment reading level. The final sample of children available for posttesting in these two treatment programs consisted of 12 dysphonetic readers and 5 dyseidetic readers. Educational validation variables chosen to measure the critical dimensions of reading accuracy and fluency included tests of single-word recognition, contextual reading, and spelling. As with the Lovett et al. (1988) study, these measures appear to have content validity for both the classification system and hypothesized responses to instruction.

Results obtained from the contextual reading measures are reported here to illustrate principles relevant to our discussion of educational validation research. First, different scores were compiled to provide a summary of average gain by subtype (dysphonetic, dyseidetic) within each treatment (ITA, Distar). Given the small number of subjects per cell and the unreliability of difference scores (Cronbach & Snow, 1977), tests of significance were not computed. Rather, graphing and visual inspection of the data were used to refine and generate hypotheses for future studies with a larger sample.

Visual inspection of average gains on measures of contextual reading suggested that, as hypothesized, dysphonetic (accuracy-disabled) children made greater gains via the language experience, analytic phonics program using the ITA than in Distar. Unexpectedly, dyseidetics in ITA also made greater gains than in the Distar condition.

Despite difficulties in sample size and assignment to treatments, the Flynn study does provide additional support for the use of direct measurements of reading and spelling behaviors in the identification of subtypes and in conducting ecologically relevant educational validation studies. Two findings may be especially important in designing future studies. First, the variability of results within each treatment condition and subtype demonstrates the inadequacy of a design that implies a simple match between reading subtype and optimal treatment program. Further, second-year data derived from the Flynn study suggest that longitudinal investigations are needed in order to measure adequately the critical dimension of reading fluency. Specifically, while only three of the eight children (38%) in the ITA condition could read fluently at grade level by the end of the first year of remediation, 9 out of 15 ITA-instructed children (60%) were fluent readers by the end of the second year.

This overview of the Lovett and Flynn studies may be helpful in designing future educational validation research. For instance, all subtypes in both investigations made the most progress in analytic phonics

programs (Lovett's DS and Flynn's ITA) and in language experience approaches that emphasized reading- and writing-connected discourse (Lovett's OWLS and Flynn's ITA). This finding suggests that subtype differentiation is unnecessary if broadly based treatment conditions are implemented. However, apart from developmental considerations that make both accuracy and fluency data important in identifying poor readers, subtyping may be important in explaining individual outcomes. That is, similar treatment outcomes may have occurred through subtle differences in teacher–child interactions within a program. This could be especially true in the language experience approaches (OWLS and ITA) but could have also occurred in the analytic phonics (DS) treatment as teachers adjusted rate and manner of presentation to each child's attentional, motivational, processing, and prior knowledge characteristics.

The possibility that these subtle adaptations were made by teachers points out the need for more dynamic aptitude × treatment interaction studies. The contamination of age ranges and concomitant and/or prior interventions in both the Lovett and Flynn studies further suggest the need for longitudinal investigations that begin in preschool and follow the children for a sufficient period of time to measure reading mastery adequately. The challenge of these and other language-based intervention studies will be to allow the teacher freedom to vary the approach and methodology while accounting for teacher–child–content–context variables that result in specific educational outcomes.

Conclusions

Identifying subtype × treatment method interactions may be one useful procedure to establish the external validity of a particular subtype solution. Findings derived from educational validation studies also serve to establish the predictive and clinical validity of a classification system. It is important to note, however, that educational validations are extraordinarily complex and difficult to conduct, primarily because of the dynamic nature of the teaching process. Proficient teachers and clinicians constantly attempt to manipulate and control learner, task, and setting (e.g., classroom) variables. Within this context, expert teachers continuously modify representations of concepts for different students, monitor and adjust the number of conceptual elements being presented, routinely induce strategies for learning, motivate students to persist in difficult-to-learn tasks, and perform all of these instructional elements simultaneously or in rapid succession to produce learning (Lyon & Moats, 1988).

Because of this complexity, our initial attempts at predicting treatment options on the basis of subtype characteristics fall far short of the

educational task confronting us. Ultimately, the best classification (subtyping) for educational purposes may result from documenting how children at risk for educational failure respond to specific instructional methodologies. Students' responses to different representations and strategies, etc. may be used as attribute variables that would allow for assignment to a particular subtype. As one example, it could be hypothesized that children with general and/or nonspecific reading delays on first grade entrance would respond equally well to any intervention methodology, whereas children at risk for failure in accuracy or fluency stages of reading would require longer periods of intervention and more specific instructional methodologies. Likewise, children who respond to reading approaches that provide explicit representations and explanations of phonological codes may differ substantially from students who benefit more from global orthographic representations. What is clear is that the appropriate use of subtyping methodology in educational settings will require creative research methodologies that are as dynamic, fluid, and flexible as is the teaching process itself.

Acknowledgments

The preparation of this chapter was made possible by a Research Scientist grant to G. Reid Lyon from the Gundersen Medical Foundation, LaCrosse, WI, and by a grant to Jane M. Flynn from the Initial Teaching Alphabet Foundation, Roslyn Heights, NY. The authors wish to thank Tanya Prindle and Mary Vaassen for critical comments on drafts of this chapter.

References

Benson, D. F., & Geschwind, N. (1975). The alexias. In D. J. Vinkis & G. W. Bruyn (Eds.), *Handbook of clinical neurology* (Vol. 4, pp. 325-350.). New York: American Elsevier.

Boder, E., & Jarrico, S. (1982). *The Boder Test of Reading-Spelling Patterns: A diagnostic screening text for subtypes of reading disability*. New York: Grune & Stratton.

Bracht, G. H. (1970). Experimental factors related to aptitude treatment interactions. *Review of Educational Research, 40*, 627-645.

Cronbach, L. J. (1977). How can instruction be adapted to individual differences? In R. M. Gagne (Ed.), *Learning and individual differences*, (pp. 79-90). Columbus, OH: Charles E. Merrill.

Cronbach, L. J., & Snow, R. E. (1977). *Aptitudes and instructional methods*. New York: Irvington.

Dunn, L. M., & Markwardt, F. D. (1970). *Manual, Peabody Individual Achievement Test*. Circle Pines, MN: American Guidance Service.

Engelmann, S., & Bruner, E (1984). *Distar Reading*. Chicago: Science Research Associates.

Fletcher, J. M., & Morris, R. (1986). Classification of disabled learners: Beyond exclusionary

definitions. In S. J. Ceci (Ed.), *Handbook of cognitive, social, and neuropsychological aspects of learning disabilities* (Vol. 1, pp. 55–80). Hillsdale, NJ: Lawrence Erlbaum.

Flynn, J. M., & Deering, W. M. (1989). Subtypes of dyslexia: Investigation of Boder's system using quantitative neurophysiology. *Developmental Medicine and Child Neurology, 31*, 215–223.

Hooper, S. R., & Willis, W. G. (1989). *Learning disability subtyping: Neuropsychological foundations, conceptual models, and issues in clinical differentiation*. New York: Springer-Verlag.

Jastak, S., & Wilkinson, G. (1984). *The Wide Range Achievement Test—Revised*. Wilmington, DE: Jastak Associates.

Lovett, M. W., Ransby, M. J., & Barron, R. W. (1988). Treatment, subtype, and word type effects in dyslexic children's response to remediation. *Brain and Language, 34*, 328–349.

Luria, A. R. (1966). *Higher cortical functions in man*. New York: Basic Books.

Luria, A. R. (1973). *The working brain: An introduction to neuropsychology*. New York: Basic Books.

Lyon, G. R. (1983). Subgroups of learning disabled readers: Clinical and empirical identification. In H. R. Myklebust (Ed..), *Progress in learning disabilities*. (Vol. 5, pp. 103–134). New York: Grune & Stratton.

Lyon, G. R. (1985a). Educational validation of learning disability subtypes. In B. P. Rourke (Ed.), *Neuropsychology of learning disabilities: Essential of subtype analysis* (pp. 228–256). New York: Guilford Press.

Lyon, G. R. (1985b). Identification and remediation of learning disability subtypes: Preliminary findings. *Learning Disabilities Focus, 1*, 21–35.

Lyon, G. R., & Flynn, J. M. (1990). Assessing subtypes of learning abilities. In H. L. Swanson (Ed.), *Handbook on the assessment of learning disabilities: Theory, research, and practice* (pp. 59–74). San Diego: College-Hill Press.

Lyon, G. R., & Moats, L. C. (1988). Critical issues in the instruction of the learning disabled. *Journal of Consulting and Clinical Psychology, 56*, 830–835.

Lyon, G. R., Moats, L. C., & Flynn, J. M. (1988). From assessment to treatment: Linkages to interventions with children. In M. Tramontana & S. Hooper (Eds.), *Issues in child neuropsychology: From assessment to treatment* (pp. 113–142). New York: Plenum Press.

Lyon, G. R., Rietta, S., Watson, B., Porch, B., & Rhodes, J. (1981). Selected linguistic and perceptual abilities of empirically derived subgroups of learning disabled readers. *Journal of School Psychology, 19*, 152–166.

Lyon, G. R., Stewart, N., & Freedman, D. (1982). Neuropsychological characteristics of empirically derived subgroups of learning disabled readers. *Journal of Clinical Neuropsychology, 4*, 343–365.

Lyon, G. R., & Watson, B. (1981). Empirically derived subgroups of learning disabled readers: Diagnostic characteristics. *Journal of Learning Disabilities, 14*, 256–261.

Rourke, B. P. (1985). *Neuropsychology of learning disabilities: Essentials of subtype analysis*. New York: Guilford Press.

Satz, P., & Morris, R. (1981). Learning disability subtypes: A review. In F. H. Pirozzolo & M. C. Wittrock (Eds.), *Neuropsychological and cognitive processes in reading* (pp. 109–141). New York: Academic Press.

Traub, M., & Bloom, F. (1975). *Recipe for reading*. Cambridge, MA: Educators Publishing Service.

Voeller, K. S., & Armus, J. (1988, November). *Analysis of word characteristics in specific reading tests, based on information-processing models: Implications for evaluating dyslexics*. Paper presented at the 37th Annual Orton Dyslexia Society conference, Tampa, FL.

Woodcock, R. (1973). *Woodcock Reading Mastery Test*. Circle Pines, MN: American Guidance Service.